图书情报工作
第63卷 第3期 2019年2月

欧洲科研开放获取基础设施项目 OpenAIRE 的建设与启示

■ 孙茜
陕西师范大学图书馆 西安 710119

摘要：[目的/意义]对欧洲科研开放获取基础设施项目 OpenAIRE 进行研究，总结成功建设经验，为国内开放获取知识库建设与发展提供参考。[方法/过程]使用文献调研和网络调研方法，对其建设模式和服务功能进行调研与分析，详细介绍其组织模式、数据源类型、互操作指导原则、数据模型、数据监管功能、检索功能、统计功能以及面向多方需求的服务功能。[结果/结论]认为其面向开放科学的内容建设模式，包括重视研究数据的开放、建设一个全球开放获取知识库合作网络平台，其支持开放获取政策执行与监管的服务功能以及自上而下、统一管理、分工合作、协同共建的管理模式值得学习借鉴。

关键词：开放获取知识库 欧洲科研开放获取基础设施 开放获取
分类号：G251
DOI：10.13266/j.issn.0252-3116.2019.03.018

1 引言

经过近20年的努力，开放获取运动取得了明显的进展，根据报道，全球论文出版后可被开放获取的份额已从2014年的18%增长至2016年的25%[1]，开放学术资源正逐步逼近“成为学术研究主流资源”[2]。今天，开放已成为一种全球共识，从开放论文到开放数据再到开放科学，这场以知识共享为特征的开放运动不断向纵深发展。

欧洲一直走在开放获取运动的前沿。2006年，欧洲研究理事会(European Research Council，ERC)发布开放获取声明[3]，2007年，提出开放获取指导原则[4]。2008年，欧盟委员会(European Commission，EC)启动开放获取先导计划，要求得到第7框架计划(7 th Framework Programme，FP7)资助的7个领域(能源、环境、健康、信息与通讯、科研基础设施、社会中的科学、社会经济学与人文科学)的资助受益人必须将其受资助产出的同行评审出版物在出版后6-12个月内提供开放获取[5]。

在此背景下，2009年，EC资助了欧洲科研开放获取基础设施 OpenAIRE(Open Access Infrastructure for Research in Europe)项目，作为支持开放获取先导计划实施的配套设施，旨在帮助科研人员完成开放获取政策要求，帮助科研管理及决策人员了解项目产出成果，监控资助受益人对开放获取政策的执行状况。OpenAIRE 的建设包括两部分内容：①建设一支遍布欧洲多国的开放获取工作团队，以支持各国开放获取事务的本地化运作；②建立一个开放获取知识库集成服务门户，实现对欧洲乃至世界各地开放获取知识库内容的自动采集与集成，并提供一站式浏览、检索、获取、统计分析等多种服务功能[6]。

中国也一直在积极建设开放获取知识库。截至2018年5月，中国科学院机构知识库服务网格已存储600 686篇论文全文，其中有313 194篇可以开放获取，自上线以来累积下载量达18 155 990篇次[7]，是目前国际科研机构中最大的公共资金资助科研成果共享系统之一[8]。2015年上线的国家自然科学基金基础研究知识库公开论文全文518 524篇，3年来论文下载量达5 194 464篇次[9]。尽管已取得一定成绩，然而，目前我国在知识库建设方面尚未涉及对全国各类型科研及教育类机构知识库内容的收集与集成，在知识库互操作及提供增值服务方面仍有较大提升空间。

笔者使用文献调研及网络调研方法，以 OpenAIRE 为关键词检索中、英文文献，并从 OpenAIRE 网站获取项目建设相关资源，通过对网站检索、浏览、统计等功能的使用，对项目的建设模式及服务功能展开全面调研。

通过文献调研发现，国外学者主要侧重从系统架构、数据模型、数据监管等技术角度对集成服务门户的

作者简介：孙茜(ORCID:0000-0002-7950-6447)，副研究馆员，硕士，E-mail:sunq@snnu.edu.cn。
收稿日期:2018-05-28 修回日期:2018-08-09 本文起止页码:138-148 本文责任编辑:徐健

建设展开研究,也有学者探讨了在科研基础设施建设背景下,研究型图书馆所应承担的职责[10]。国内学者聂华通过对 OpenAIRE-CORE2014 年会的观察与思考,介绍了项目的建设背景、目标及现状[11];王翠萍、王佳佳将 OpenAIRE 集成服务门户视为科研数据知识库注册目录系统,从资源收录量、检索功能、服务模式等维度对其进行调研分析[12]。

已有研究尚未有从项目的组织与管理模式、内容建设模式及服务功能等多个维度对其成功建设经验进行分析与总结,本文研究期望能为国内开放获取知识库的建设与发展提供参考借鉴。

2 OpenAIRE 的组织模式

2.1 OpenAIRE 的组织架构

尽管开放获取是一项全球性事业,然而其实施却离不开本地化运作。此外,开放获取的深入推进会同时受到政策、法律、管理、服务、技术等多种因素的共同影响。基于上述认知,OpenAIRE 项目汇聚了来自欧洲多国高校、研究型图书馆、研究中心、开放学术组织等多类型机构的信息技术专家、法律专家、数据专家与图书馆员,通过协同共建的方式实现项目总体目标。

为保障对所有合作机构的有序管理,针对建设目标,项目建立起项目协调人、项目经理、区域管理团队、技术管理团队这样一种分级、分类的组织架构。项目协调人负责项目的整体协调与管理,由 4 名人员组成,其中全局协调员 2 名,科学、技术协调员各 1 名。项目经理负责项目的组织与实施,由 5 名人员组成,其中项目经理 1 名、科学经理 3 名、技术经理 1 名。区域管理团队负责服务人员团队的组建与运作,以支持 EC 开放获取政策在各国落地执行,由东欧、西欧、南欧、北欧 4 个区域管理办公室及 33 个国家(28 个欧盟成员国及挪威、冰岛、克罗地亚、瑞士及土耳其)的国家开放获取办事处(National Open Access Desks,NOAD)组成。技术管理团队负责数字基础设施的建设,包括 OpenAIRE 集成服务门户的运行与维护、数据的集成与处理,由软件工程师、数据工程师、数据架构师等 18 名技术负责人组成[13]。

2.2 OpenAIRE 的管理模式

在上述组织架构的基础上,项目采纳自上而下、统一管理、分工合作、协同共建的管理模式,根据项目内容的专业性质首先将项目划分为若干子项目,每个子项目由一所专门机构负责,然后将子项目分解为若干项工作任务,每项任务再交由一所主管机构带领几所参与机构共同组织实施,通过分工合作的方式协同完成工作任务。表 1 以 OpenAIRE 三期项目 OpenAIRE2020 为例,展示其下设的 10 个子项目、子项目内容与子项目主管机构[14]。可以看出,这种管理模式能够充分发挥各个合作机构的专业优势。

表 1 OpenAIRE2020 项目下设子项目、子项目内容及主管机构

序号	子项目	子项目内容	子项目主管机构
1	协调与探索	负责项目的总体协调与管理;探索项目的可持续发展	雅典大学
2	推广与宣传	制定宣传计划、制作宣传材料、寻找宣传渠道、组织宣传活动;负责提高项目可见度与认知度	哥廷根大学
3	与国际其他地区的开放获取知识库结盟	与世界其他地区的开放存取知识库建立联盟,促进彼此之间的互操作。	开放获取知识库联盟
4	支持与培训	对 NOAD 进行组织与管理;面向科研人员、科研管理人员及知识库管理员分别开展培训活动;负责网络咨询台的实施与运作	米尼奥大学
5	开放获取论文出版	研究出版开放获取论文的工作流、方法与费用	欧洲研究图书馆联盟
6	OpenAIRE 平台运行与维护	负责 OpenAIRE 平台的管理与监控;负责软件部署与升级;进行系统备份与恢复	华沙大学数学与计算建模跨学科中心
7	调研学术交流的未来	研究数据隐私保护问题;研究评价开放获取影响力的指标;研究建立文献与其引用数据的关联;研究学术论文的开放评审	哥廷根大学
8	信息空间维护与扩充	开发并维护数据模型;增补元数据;管理数据流;发布关联开放数据	意大利国家研究委员会信息科学与技术研究所
9	前端服务	通过 OpenAIRE 集成服务门户提供统计、汇总、可视化等功能;提供 API 供第三方获取数据	雅典研究与创新中心
10	知识抽取	研发信息推理、聚类、分类、实体解析与关联以及构建学术交流网络的算法	华沙大学数学与计算建模跨学科中心

2.3 OpenAIRE 的服务模式

作为支持 EC 开放获取政策执行与监控的配套设施,OpenAIRE 的服务模式是通过分布各国的 NOAD 帮助本国科研人员将受资助产出的出版物及研究数据存

储到本人所在机构或合适的主题知识库之中，通过 OpenAIRE 集成服务门户对这些开放获取知识库元数据的定期自动收割与集成，提供统一的检索、浏览、获取、利用入口，从而实现帮助科研人员完成开放获取强制性政策要求。

作为当地的开放获取专家，NOAD 深入本国科研机构，主动与科研及管理人员沟通交流、了解需求、宣传政策、解答咨询、解决问题，小到帮助科研人员解决自存储过程中碰到的技术与法律问题，大到向科研机构或政府决策者提供开放获取政策建议，其具体承担的职责包括：宣传推广开放获取理念、政策及最佳实践；解答科研人员自存储过程中遇到的各种技术、法律问题，帮助他们完成自存储操作；帮助开放获取知识库管理员了解 OpenAIRE 提出的系统互操作指导原则，确保元数据可以被自动收割；协助各类管理人员对科研项目成果进行统计分析；解答有关 EC 开放获取政策的咨询，向各级决策者提供制定本地开放获取政策的建议[15]。

3 OpenAIRE 的内容建设模式

3.1 数据源类型

OpenAIRE 致力于收集如下 3 类科研成果：①全球所有经过同行评议的开放获取出版物，包括期刊论文、会议论文、学位论文、研究报告、图书（章节）等，这些出版物的元数据中最好应包含项目及资助信息；②各种国家级资助机构资助产出的科研出版物，无论这些出版物是否可以开放获取；③受资助科研项目产出的研究数据或者是被 OpenAIRE 集成服务门户中出版物引用的研究数据[16]。为了获得上述内容，OpenAIRE 从一系列数据源那里采集数据，各类型数据源及所提供内容如表 2 所示：

表 2 OpenAIRE 数据源类型及提供内容

数据源类型	提供内容
开放获取文献知识库	各类型开放获取出版物信息
开放获取数据知识库	各类型开放获取研究数据信息
科研管理系统	科研成果、机构、项目、人员信息
实体注册目录系统	注册实体的经过验证的可靠内容，如项目、机构、人员等实体信息
开放获取知识库集成门户	开放获取出版物及研究数据信息

其中，开放获取文献知识库包括机构知识库、主题知识库，许多大学机构知识库、知名的 Europe PMC[17] 以及美国康奈尔大学图书馆维护的主题知识库 arXiv[18] 属于此类型数据源。开放获取数据知识库提供研究数据、软件应用程序、视听材料等多种类型研究数据，Dryad[19]、Figshare[20] 属于此类型数据源。科研管理系统不仅提供研究成果、机构、项目、人员信息，还提供它们之间的语义关联信息。实体注册目录系统提供注册的项目、机构、人员等经过验证的可靠信息，可被用来补充、验证 OpenAIRE 集成服务门户中已有信息。例如，从 EC 研究与发展信息服务系统（Community Research & Development Information Service，CORDIS）[21] 获取 FP7 和地平线 2020（Horizon2020，H2020）框架资助计划资助的项目名称、资助金额、开始时间、结束时间、内容简介、项目协调人、负责机构等具体信息；从开放研究者与贡献者标识符系统（Open Researcher & Contributor Identifier，ORCID）[22] 获取全球科研工作者的成果、项目等信息；从开放获取知识库登记目录（Open Directory of Open Access Repositories，OpenDOAR）[23] 和研究数据知识库登记目录（Registry of Research Data Repositories，re3data）[24] 获取开放获取知识库的名称、类型等信息。利用这些经过验证的可靠信息，可以为出版物或研究数据等研究成果元数据补充项目、机构、作者等情境信息。开放获取知识库集成门户通过对多个知识库元数据进行自动收割与集成，支持从统一的入口实现对多个知识库内数据的统一检索与获取利用，包括机构知识库集成门户、开放获取期刊集成门户、研究数据知识库集成门户等。拉丁美洲 La Referencia 知识库[25]、日本机构知识库在线（Japanese Institutional Repositories Online，JAIRO）[26]、开放获取期刊目录（Directory of Open Access journals，DOAJ）[27] 及开放获取数据知识库集成门户 DataCite[28] 都属此类型数据源。

3.2 互操作指导原则

为了实现与不同类型数据源的数据交互，实现对元数据的自动收割，OpenAIRE 目前针对 3 类不同数据源分别制定了 3 套互操作指导原则，分别是《OpenAIRE 对文献知识库的指导原则》、《OpenAIRE 对数据知识库的指导原则》以及《OpenAIRE 对基于 CERIF-XML 数据格式科研管理系统的指导原则》[29]。

在这些指导原则里，OpenAIRE 指明了建议使用的元数据规范及数据传输协议。在元数据规范方面，OpenAIRE 分别针对不同数据源规定了应遵循的元数据规范：文献知识库使用都柏林核心（Dublin Core，DC）元数据规范；数据知识库使用 DataCite 元数据规范；科研管理系统使用通用欧洲科研信息格式（Common European Research Information Format，CERIF）数

据模型,并基于 CERIF - XML 数据交换格式与 OpenAIRE 集成服务门户进行数据交互。在传输协议方面,要求各类数据源均使用开放文档先导 – 元数据收割协议(Open Access Initiative-Protocol Metadata Harvesting,OAI-PMH)协议。

为满足对开放获取政策执行状况的监控需求,成果元数据中需要具备项目、资助、使用权限及获取时滞信息,因此,《OpenAIRE 对文献知识库的指导原则》里还要求在出版物元数据中应尽可能包含如下要素:①资助主体及项目编号;②权限信息及获取模式;③获取时滞;④出版物及研究数据的永久标识符[30]。

3.3 数据模型

OpenAIRE 集成服务门户不仅对出版物、研究数据进行建模,同时还对项目、资助、机构、数据源进行建模,使系统同时容纳了科研成果及科研情境信息,并建立起它们之间的关联关系,这一方面有助于了解科研成果产生的情境源流,另一方面也是统计资助机构资助科研项目产出科研成果的必备数据基础。

由于需要容纳出版物、研究数据、项目、资助、机构、数据源这些实体数据,而且要容纳上述实体之间的关联关系,这就对 OpenAIRE 集成服务门户数据模型的灵活性与可扩展性提出了较高要求,一方面需要灵活容纳新实体并可随时增加实体之间新的关系,另一方面,在添加新数据的时候,不应影响系统中原有的信息模式。

OpenAIRE 集成服务门户选择 CERIF 作为其底层数据模型,该模型创新性的实体建模方法包括将基础实体与关联实体相分离,使用关联实体表达实体间关系,使用语义层统一存储、维护各种术语并实现被关联实体的调用[31]。这种建模方法赋予 CERIF 按照需求灵活引入新实体、表达新属性及实体间新关系的能力,满足了 OpenAIRE 集成服务门户数据建模的需求。

OpenAIRE 首先确定了数据源(data source)、科研成果(result)、人员(person)、机构(organization)、项目(project)及资助(funding)等基础实体。然后,使用关联实体表达实体间的关联关系,通过项目与资助(project_funding)、项目与成果(project_result)、项目与机构(project_organization)、人员与成果(person_result)、机构与成果(organization_result)、机构与资助(organization_funding)等关联实体,支持捕获资助机构资助了哪些项目、项目产出了哪些成果、成果归属于哪些人员和机构等信息,如图 1 所示。OpenAIRE 使用语义层统一存储并维护表达实体间关联关系的语义或实体属性取值的受控词表,如定义组成(is part of)这个词汇表达各级基金资助计划之间的等级关系,用以描述资助机构分项资助计划与总资助计划之间的关系,为统计资助机构各级资助计划产出的科研成果奠定数据基础[32]。图 2 展示了 FP7 下属 4 个专项计划——合作计划(Cooperation)、原始创新计划(Ideas)、人力资源计划(People)、研究能力建设计划(Capacity)及 1 个核研究特殊计划——欧洲原子能共同体计划(Euratom)与 FP7 总体框架资助计划之间的关系。

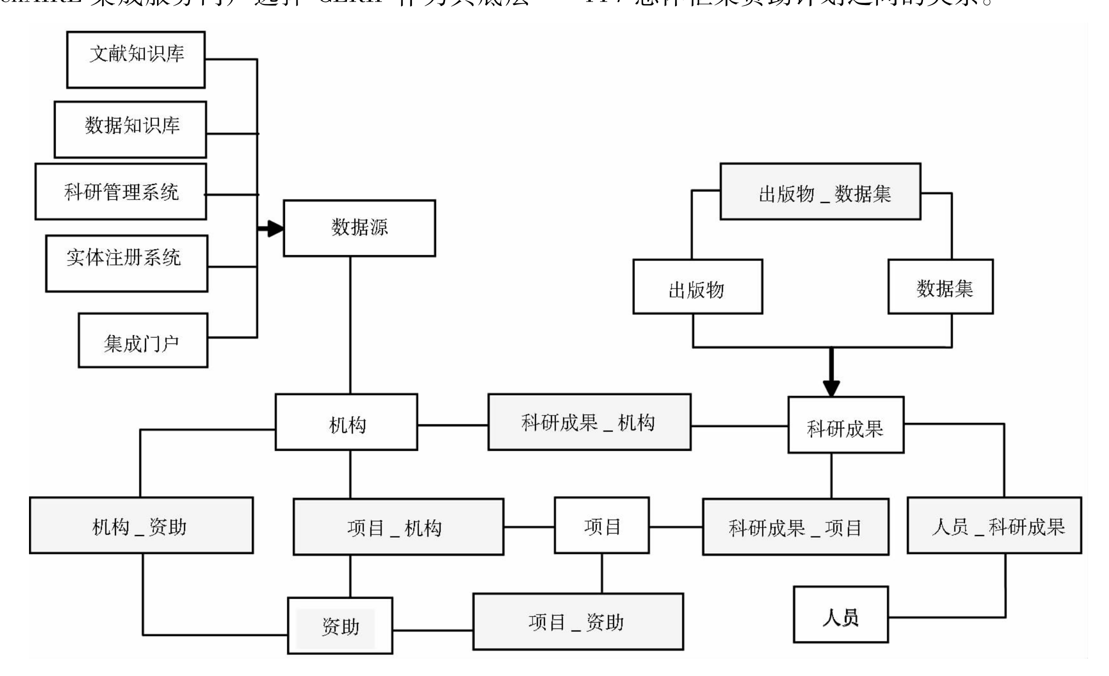


图 1　OpenAIRE 集成服务门户实体模型

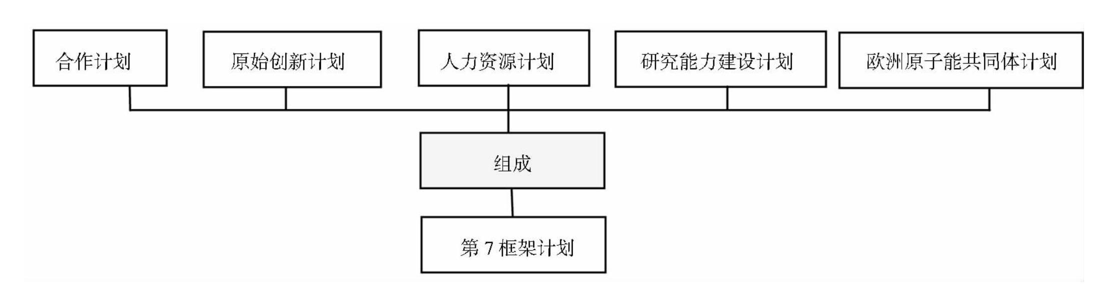


图2 FP7分项资助计划与总资助计划关系

尽管不同数据源的元数据格式各不相同,有DC格式,也有DataCite格式元数据,但OpenAIRE集成服务门户数据模型中完备的实体及实体间关系却可以完整容纳所有数据源数据,使它们分别映射到模型中对应的基础实体与关联实体,建立起资助计划与受资助项目及项目产出成果的关联关系,从而奠定了用于发现受资助科研机构、科研人员执行科研成果开放获取政策情况的数据基础,进而可支持实现监控开放获取政策执行效果的科研管理功能[33]。

4 服务功能

4.1 数据监管功能

作为支持EC开放获取政策的重要基础设施,数据的准确完整性至关重要,它直接影响统计结果的正确性,因此数据监管在OpenAIRE集成服务门户建设中具有高度重要的地位。

数据监管指从数据产生开始,对数据进行管理以促进利用,以保证数据符合使用目的和要求,且可用于知识发现和再利用。对于动态数据来说,这意味着持续的更新,以使其符合使用目的。较高层次的数据监管还包括维护已出版资料的注释链接和其它链接[34]。

OpenAIRE集成服务门户使用D-NET软件管理平台,该软件平台提供了一套丰富的数据管理功能,支持分层的数据集成管理功能,可连接起通过多种途径进入系统的多种类型的来源数据[35]。OpenAIRE集成服务门户的数据流程处理如图3所示:

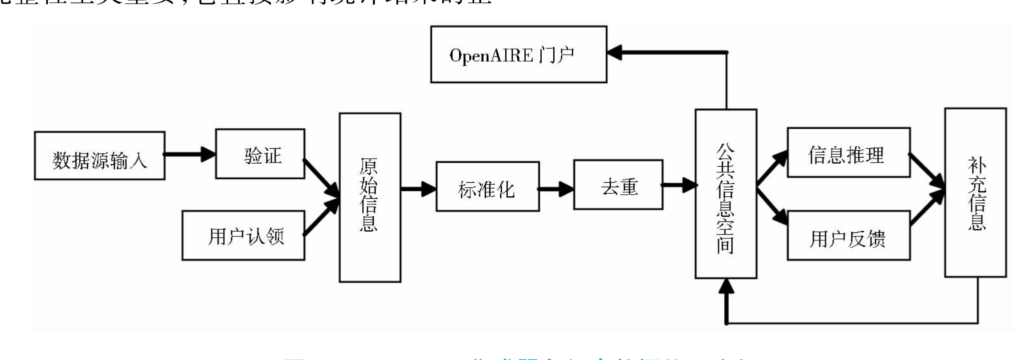


图3 OpenAIRE集成服务门户数据处理流程

OpenAIRE集成服务门户通过两种途径获取原始信息:①从通过验证的数据源那里自动收割元数据。开放获取知识库管理员在提交数据前,需要运行OpenAIRE集成服务门户提供的验证程序来检查其元数据是否可以被自动收割,并初步判断元数据质量,通过验证的知识库可以向OpenAIRE集成服务门户提交内容。即使数据源已经通过验证,开始提交数据,由于后期软件升级、元数据标准改变、地址变更等多种因素都会影响自动收割元数据过程,因此需要对数据源进行定期验证。②在集成服务门户网站上为用户提供项目成果认领功能,即用户可以为出版物、研究数据的元数据添加关联的项目、资助等信息,也可为项目元数据添加产出的出版物及研究数据等信息。

通过上述两种途径采集的原始信息往往存在着数据不完整、重复甚至错误等各种问题,这时需要进行数据清洗,数据清洗包括格式转换、补充必要信息、去重等步骤。由于各个数据源的元数据格式彼此不同,有DC格式、也有DataCite格式等,因此需要把它们统一转换为CERIF数据格式。此时还需要对原始数据源的信息补充一些必要信息。例如,需要为每条数据添加数据来源,这一方面是为了实现数据源的可见性,另一方面也是为了当数据出现问题时,方便溯源查找。来自于多个数据源的元数据通常有所重复,需要对这些数据进行合并去重,OpenAIRE集成服务门户基于数

据的相似度比较算法来判断记录是否重复。

清洗后的数据可以初步投入使用,然而这些数据需要进一步补充与丰富。由于原始元数据中往往缺乏项目、资助、使用权限及获取模式等信息,而这些内容是实现监控开放获取政策执行状况不可或缺的,因此,OpenAIRE 集成服务门户运用数据挖掘及信息推理技术对全文及元数据进行分析,例如,从全文致谢部分发现并提取项目及资助信息;从全文参考文献中发现引用研究数据的持久标识符,从而建立起出版物与研究数据之间的关联;利用文献自动分类技术对文献进行初步分类,自动生成主题词;利用聚类技术对文献进行聚类,形成相似文献,建立起文献之间的关联[36]。这些经过信息推理算法产生的内容不仅可以补充原始数据中所缺乏的项目、资助等必备内容,其余更多丰富内容还可以为提供其它增值服务奠定基础。此外,门户网站也接受用户对元数据的反馈意见,并根据反馈意见对元数据继续进行完善。

4.2 检索服务功能

面对科研活动中利益相关各方获取多种科研信息的需求,OpenAIRE 集成服务门户除了提供出版物、研究数据等科研成果信息之外,还提供项目、机构、数据源等科研情境信息。为帮助用户便捷发现所需信息,对出版物、研究数据、项目、机构、数据源这 5 种对象均提供了多种检索及浏览途径,如表 3 所示。检索及浏览途径都比较丰富,这对于在一个数据容量巨大、内容类型丰富、语言多样、来源广泛的信息空间中快速、准确定位查询对象是非常必要的。

发现查询对象后,用户可以在各对象页面了解其基本属性信息,OpenAIRE 集成服务门户还在每个对象页面关联展示相关的其它对象,例如在成果(出版物、研究数据)页面关联展示成果所属项目、在项目页面关联展示项目产出成果、在机构页面关联展示机构所负责的项目或所管理的数据源、在数据源页面关联展示该数据源所提供的成果。这种在一个对象页面关联展示相关其它对象的方法,有助于进一步直接发现相关信息,便于管理人员了解项目产出全部成果,也便于研究人员通过更多成果及情境信息掌握更全面的科研过程。各个对象页面展示的主要属性信息、关联对象及其它信息如表 3 所示。值得指出的是,为满足项目负责人了解项目进展、汇报项目成果的需求,项目页面提供了可自动生成项目报告、建立出版物列表的应用程序[37]。

表 3 OpenAIRE 集成服务门户检索服务功能

对象	检索途径	过滤条件	浏览途径	页面展示内容
出版物	5 种(题名、作者、出版者、主题、全部字段)	9 种(文献类型、语言、资助机构、资助计划、一级资助类别、二级资助类别、出版时间、获取模式、数据源)	7 种(资助机构、获取模式、出版时间、文献类型、语言、数据源、团体)	(1)展示出版物题名、作者、类型、来源、主题、持久标识符等基本属性信息;(2)关联展示引用出版物及研究数据、受资助项目、数据源
研究数据	5 种(题名、作者、出版者、主题、所有字段)	6 种(资助机构、获取模式、出版时间、内容类型、语言、数据源)	6 种(资助机构、获取模式、出版时间、内容类型、语言、数据源)	(1)展示研究数据名称、出版者、类型、简介、持久标识符等基本属性信息;(2)关联展示施引出版物、受资助项目、数据源
项目	4 种(项目首字母缩写名称、项目名称、关键词、所有字段)	6 种(资助机构、资助计划、一级资助类别、开始时间、结束时间、39 号特别条款)	3 种(资助机构、开始时间、结束时间)	(1)展示项目名称、资助机构、项目编号、开始、结束日期、责任机构;(2)关联展示项目产出出版物、研究数据;(3)展示项目产出成果的统计信息;(4)提供应用程序,支持生成项目产出出版物列表、项目报告等
机构	3 种(名称、缩写、所有字段)	2 种(国家、类型)	2 种(国家、类型)	(1)展示机构名称、国家;(2)关联展示机构负责的科研项目及管理的数据源
数据源	4 种(名称、英文名称、主题、所有字段)	4 种(数据源类型、语言、内容类型、OpenAIRE 兼容性版本)	4 种(数据源类型、语言、内容类型、OpenAIRE 兼容性版本)	(1)展示数据源名称、类型、提供的内容、OpenAIRE 兼容性版本等基本属性信息;(2)关联展示提供的出版物、研究数据

4.3 统计服务功能

以包含了成果、项目、资助、获取模式等丰富内容的元数据为基础,OpenAIRE 集成服务门户实现了对科研项目产出成果的统计功能,可以从成果数量、出版时间、获取模式等多种维度进行统计。

图 4 – 图 7 分别展示了 FP7 及下属分项资助计划资助的科研项目产出成果的部分统计数据。其中,图 4 展示了 FP7 下属 4 个分项资助计划及 1 个核研究特殊计划资助产出的科研成果数量,从图 4 中可以看出,在 5 项资助计划中,合作计划资助项目产出的出版物数量最多,达到 109 071 篇;图 5 展示了各种获取模式下的科研成果数量,可以看到,目前有接近 70% 的成果已实现了开放获取;图 6 展示得到 FP7 资助的产出出版物数量最多的前 5 项科研项目;图 7 展示各年产

出的得到 FP7 资助的出版物数量，从图 7 中可以看出，2013 年出版的成果数量达到峰值，共出版 46 613 篇出版物。

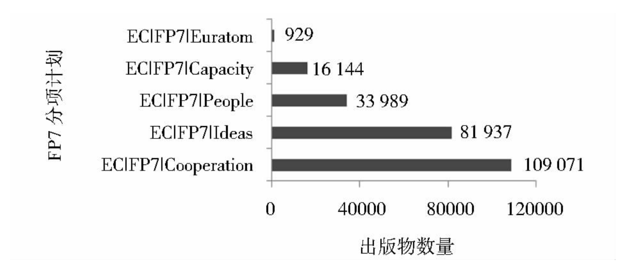


图 4 FP7 各个分项计划资助产出的出版物数量

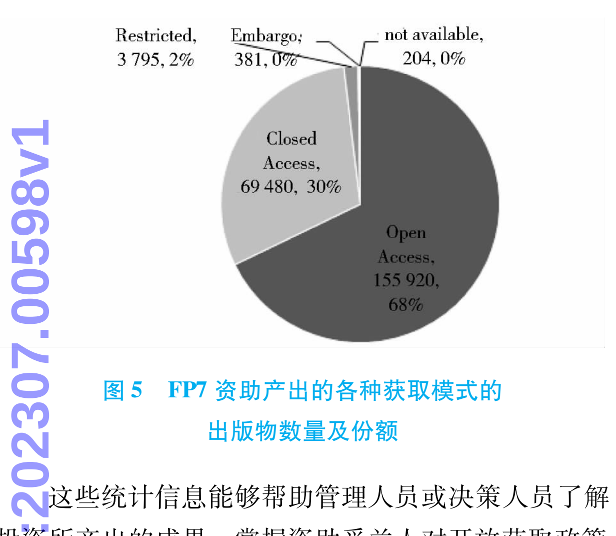


图 5 FP7 资助产出的各种获取模式的出版物数量及份额

这些统计信息能够帮助管理人员或决策人员了解投资所产出的成果、掌握资助受益人对开放获取政策

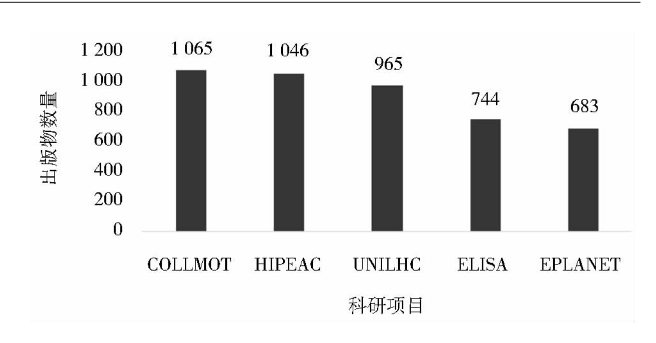


图 6 FP7 资助产出出版物数量最多的前 5 名科研项目

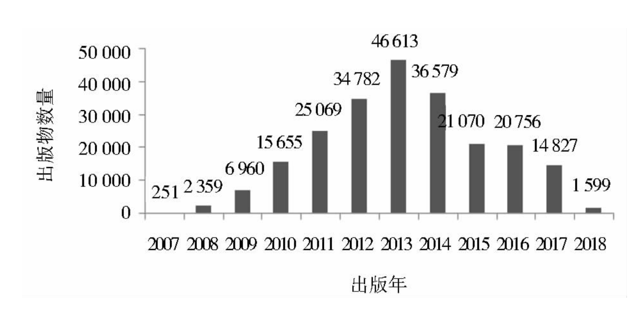


图 7 各年出版的得到 FP7 资助的出版物数量

的执行状况，为决策提供了数据支撑。

4.4 面向多方需求的服务功能

OpenAIRE 面向利益相关多方的需求，如科研人员执行开放获取强制政策、项目负责人了解项目产出、资助机构了解资助产出成果、掌握开放获取政策执行情况、数据源管理员增强显示度等各种不同需求，分别提出了相应的使用建议及对应的服务功能，如表 4 所示：

表 4 利益相关方的需求、使用建议及服务功能

利益相关方	需求	使用建议	服务功能
科研人员	(1)完成开放获取强制性政策要求；(2)获取学术资源	(1)向开放获取知识库提交科研成果，包括出版物及研究数据；(2)提交成果元数据，应包括项目、资助、权限、时滞期、持久标识符等信息；(3)确保成果开放获取	(1)自动收割开放获取知识库元数据；(2)提供统一检索、浏览、获取、使用功能
项目负责人	(1)了解项目产出成果；(2)汇报项目进展报告	认领项目产出出版物、研究数据	(1)在项目页面关联展示项目产出的出版物、研究数据及统计信息；(2)生成项目产出成果报告
资助机构	(1)了解各级各类资助计划产出科研成果；(2)掌握资助受益人对开放获取强制政策的执行情况	提供项目编号、项目名称与缩写、开始时间与结束时间、资助类别、资助计划、资助机构等信息	(1)从资助机构、资助计划、资助类别等途径浏览产出科研成果；(2)提供成果数量、出版时间、获取模式等多维统计功能
数据源管理员	(1)提交数据源数据；(2)增加数据源可见度	(1)确保知识库在 OpenDOAR 或 re3data 网站上注册；(2)自查是否符合互操作指导原则；(3)运行验证程序，检查元数据是否可被自动收割	(1)提供互操作指导原则及验证程序；(2)从通过验证的数据源自动收割元数据；(3)对数据源数据进行清洗、补充及完善；(4)在成果页面展示数据源信息

5 OpenAIRE 带给我们的启示

OpenAIRE 的建设对于欧洲开放获取的深入发展起到了积极的推动作用，不仅推动欧洲多国制定和出台了多项开放获取政策，在数字基础设施建设方面也卓有成效，除了有效推进各国开放获取知识库的建设与普及之外，OpenAIRE 集成服务门户也以其丰富的科研信息成为获取全球开放学术资源的重要学术交流平台，以其完善、便捷的服务功能成为支持欧洲开放获取政策执行与监控的重要服务平台。其成功建设带给我们如下启示：

5.1 面向开放科学的内容建设模式

(1)重视研究数据的开放获取。开放科学坚守科学的基本信条：所有的理论都应该经得起批评，而且结

论应可重复。开放科学坚持科学必须开放且可重复,研究的所有要素均应对外公开。研究数据可用于验证研究成果,开放科学重视研究数据的开放,强调数据的再利用,把研究数据置于和研究论文同等重要的地位,不仅倡导论文的开放获取,认为论文中数据的开放也是不可缺少的[38]。

研究数据的发现与重用离不开情境信息的支持,根据研究数据得到的项目支持、施引论文、责任机构等信息可以帮助使用者判断数据的潜在使用价值。OpenAIRE 集成服务门户重视对科研成果和情境信息的收集,从多类型数据源中获取科研信息,面对许多成果元数据缺失必备信息的现实问题,使用数据挖掘和信息推理技术从全文及元数据中推理出这些内容。通过对元数据进行补充与丰富,OpenAIRE 集成服务门户建立起出版物与研究数据的关联关系,建立起科研成果与科研情境的关联关系,这使得研究人员不仅能够获取研究数据来验证研究成果,同时还可以根据自身需求进行数据重用。重视对研究数据的开放获取,能够更好地发挥研究数据对科学研究的推动作用。

(2)建立全球开放获取知识库合作网络平台。OpenAIRE 集成服务门户致力于收集全球开放获取学术资源,通过建立一个全球开放获取知识库合作网络平台,对全世界提供开放获取资源的一站式获取,更好地满足科学研究与学术交流之需。

OpenAIRE 集成服务门户针对不同类型知识库制定了不同的互操作指导原则,以实现对元数据的自动获取;使用 CERIF 数据模型,以容纳来自多源系统的异构数据,该模型是欧盟向其成员国推荐使用的科研管理系统数据模型的官方标准,是科研信息存储与互操作的国际标准,是整合多源异构数据的通用数据模型。

截至 2018 年 5 月,OpenAIRE 集成服务门户从来自全球的 11 675 个数据源中自动收集数据,已涵盖 23 944 971条出版物记录及 696 306 条研究数据记录。如前文所述,全球许多知名的开放获取知识库均是 OpenAIRE 集成服务门户的数据源。提供出版物记录最多的前 5 位数据源如表 5 所示:

表 5　提供出版物数量最多的前 5 位数据源

数据源名称	提供出版物记录数量(条)	提供出版物数量份额
Europe PMC	4 418 523	18%
DOAJ	3 260 571	14%
JAIRO	2 258 679	9%
arXiv	1 384 263	6%
LA Referencia	1 263 863	5%

5.2　面向开放获取政策的服务定位

开放获取知识库在建设过程中面临的最大问题是可持续问题,寻求政策支持、资金资助是发展的基本保障。已有实践表明,得到政策与资金支持的开放获取知识库的建设与发展都较为顺利,如日本 JAIRO 和我国中科院机构知识库,与此相反,中国高校机构知识库联盟由于没有得到资金的长期资助,在预研项目结束之后,发展速度明显降低[39]。

与国家、资助机构强制性开放获取政策紧密合作,并提供服务已被证明是开放获取知识库可持续发展方向之一[40]。只有提供支持开放获取政策执行与监管的服务功能,满足国家与资助机构的管理需求,才可能进而得到他们的持续支持。

2008 年,EC 提出开放获取先导计划,对得到 FP7 资助的 7 个主题领域科研项目产出成果提出开放获取要求。2009 年,作为先导计划的配套支持设施,EC 投资 490 多万欧元资助建设 OpenAIRE 项目,收集上述 7 个主题领域开放获取出版物并建立成果与资助项目的关联[41]。2011 年,EC 继续投资 510 多万欧元建设 OpenAIREPlus 项目,该项目进一步扩大了收集成果的学科范围和类型,将研究数据及其它国家级资助机构的资助成果也纳入收集范围[42]。伴随着 H2020 框架计划里所有研究领域受资助成果强制性开放获取政策的颁布以及开放研究数据先导计划的启动,EC 开启了对 OpenAIRE 的第 3 轮投资,投资 1 300 万欧元启动 OpenAIRE2020 项目,用以支持 H2020 框架计划强制性开放获取政策[43]。

除了实现了对 FP7 资助科研成果的统计功能之外,目前 OpenAIRE 集成服务门户还实现了对欧洲更多国家级资助机构,如英国研究理事会(Research Council United Kingdom,RCUK)、瑞士国家科学基金会(Swiss National Science Foundation,SNSF)、奥地利科学基金会(Austrian Science Fund,FWF)等资助成果的统计功能[44-46]。除了欧洲之外,加拿大研究图书馆协会(Canadian Association of Research Libraries,CARL)于今年年初宣布与 OpenAIRE 开展重要合作,利用其服务功能来识别加拿大的开放获取内容,以使科研共同体更好地了解加拿大资助机构或科研机构产出的哪些出版物是可以开放获取的,并确保加拿大可以直接参与到快速发展的开放科学的前沿功能的讨论中[47]。

5.3　自上而下、统一管理、分工合作、协同共建的管理模式

在开放获取知识库的建设模式中,有自下而上和

自上而下两种模式,前者由于缺乏政策支持,发展容易陷入困境。而后者由于得到决策层的支持,可以对项目建设进行总体规划和全面部署,从而能够促进开放获取知识库的可持续发展[48]。

作为一项泛欧项目,人们在建设伊始即认识到,为了保证项目的顺利实施与成功运作,需要来自各个国家、多个行业人员的协同合作。OpenAIRE 项目合作机构从一期的 30 余所增长至三期的 50 所,充分展示出项目建设的合作性这一本质特征。面对众多的合作机构和广阔的地域范围,OpenAIRE 项目建立起分级、分类的组织架构,设置项目协调人与项目经理人对项目进行统一协调与组织实施,针对项目服务人员团队的组建与运作设置区域管理团队进行管理,针对项目数字基础设施的建设与运行设置技术管理团队进行管理。在这种组织架构基础之上,采纳自上而下、统一管理、分工合作、协同共建的管理模式,将项目进一步划分为子项目,再将子项目分解为工作任务,根据工作任务的专业性质交由专业机构共同合作完成。此外,项目通过遍布欧洲多国的 NOAD 深入本国科研机构为科研及管理人员提供咨询与帮助的服务模式,有效实现 EC 开放获取政策在当地的成功实施。

OpenAIRE 的组织架构、管理模式与服务模式兼具集中与分布的优点,一方面能够统一规划、组织协调,另一方面能够因地制宜、灵活部署,同时还能够最大限度地发挥合作机构的专业特长,通过参与机构的分工合作实现项目的协同共建,其建设模式被实践证明是行之有效的。

OpenAIRE 项目自建设以来,取得了一系列令人瞩目的成果。①在有效推动开放获取在欧洲多国的深入发展方面,奥地利、保加利亚、克罗地亚都新建了遵从 OpenAIRE 集成服务门户互操作指导原则的国家级开放获取知识库,卢森堡建设了国内第一个开放获取知识库,马耳他建立了国内第一个大学机构知识库,波兰开发了波兰版 DSpace,几年里建设的机构知识库数量增长了几乎 4 倍。奥地利、克罗地亚制定了学位论文开放获取政策,意大利教育部颁布了与 EC 要求一致的科研出版物及研究数据的开放获取政策,拉脱维亚起草了大学的开放获取政策,罗马尼亚在国家创新战略中要求对所有科研成果实现开放获取,瑞士科学基金委员会提出对专著的开放获取要求[49]。②在促进与世界其他地区建立开放获取知识库合作联盟方面,通过与国际开放获取知识库联盟的合作,已经成功实现与拉丁美洲最大的机构知识库 La Referencia 的互操作,未来还计划与包括中国、美国在内的世界上更多地区与国家展开合作。③在项目的自身建设与推广方面,通过一系列的学术科学研究、通过多种形式的培训与工作会议、通过各种媒体的宣传与推介活动,OpenAIRE 现在已经成为推动欧洲开放获取深入发展的关键参与者,在全球范围具有很高的影响力。

6 结语

在国内目前已建成的开放获取知识库中,重视研究数据的开放获取、支持研究数据管理且建立起研究数据与施引论文之间关联的知识库还不多见。研究数据的开放共享对科学研究的推动作用毋庸置疑,未来我们需要在开放获取知识库的建设中将研究数据纳入内容收集范围并提供研究数据管理服务功能。

借鉴 OpenAIRE 的建设成功经验,我国也应思考规划建设国家级开放获取基础设施,以我国各个教育机构或科研机构的知识库作为数据源,采用自上而下、统一管理、分布建设、集中展示的建设模式,面向机构科研管理需求设计服务功能,以支持我国开放获取政策为目标,这样或许可以突破目前我国许多地区高校机构知识库建设的困境[50],带动我国机构知识库持续深入发展,这对于推动我国开放科学发展进程也将具有积极意义。

参考文献:

[1] MICHAEL J, ANDREW P, STEPHANIE O, et al. Monitoring the transition to Open Access December 2017[EB/OL]. [2018 - 02 - 22]. http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/monitoring-transition-open-access-2017.pdf.

[2] 张晓林, 李麟, 刘细文, 等. 开放获取学术信息资源:逼近"主流化"转折点[J]. 图书情报工作, 2012, 56(9):42 - 47.

[3] MARIEKE G. Open Access to research data - timeline[EB/OL]. [2018 - 03 - 10]. https://access.okfn.org/2015/04/30/open-access-to-research-data-timeline/.

[4] ERC Scientific Council Guidelines for Open Access[EB/OL]. [2018 - 01 - 04]. https://erc.europa.eu/sites/default/files/document/file/erc_scc_guidelines_open_access.pdf.

[5] Open Access in FP7 (Seventh research framework Programme) [EB/OL]. [2018 - 01 - 15]. https://www.openaire.eu/open-access-in-fp7-seventh-research-framework-programme.

[6] OpenAIRE [EB/OL]. [2017 - 11 - 01]. https://cordis.europa.eu/project/rcn/93855_en.html.

[7] 中国科学院机构知识库服务网格[EB/OL]. [2018 - 02 - 10]. http://www.irgrid.ac.cn/.

[8] 张冬荣, 祝忠明, 李麟,等. 中国科学院机构知识库建设推广

与服务[J]. 图书情报工作, 2013, 57(1):20 – 25.
[9] 国家自然科学基金基础研究知识库[EB/OL]. [2018 – 03 – 22]. http://or. nsfc. gov. cn/.
[10] LOSSAU N. An Overview of Research Infrastructures in Europe - and recommendations to LIBER[J]. Liber quarterly the journal of European research libraries, 2012, 21(3/4):313 – 329.
[11] 聂华. 全球开放获取运动的新进展——OpenAIRE-COAR2014 年会的观察与思考[J]. 大学图书馆学报, 2015, 33(2):5 – 12.
[12] 王翠萍, 王佳佳. 科研数据知识库注册目录系统调查与分析[J]. 情报资料工作, 2017(5): 56 – 62.
[13] OpenAIRE Contact us[EB/OL]. [2018 – 08 – 05]. https://www. openaire. eu/contact-us.
[14] OpenAIRE 2020 Work Packages, Tasks, Deliverables[EB/OL]. [2018 – 08 – 05]. https://www. openaire. eu/openaire2020-wps-tasks.
[15] RETTBERG N, SCHMIDT B. OpenAIRE - building a collaborative open access infrastructure for European researchers[J]. Liber quarterly the journal of European research libraries, 2012, 22(3): 160 – 175.
[16] Factsheet for repository managers[EB/OL]. [2018 – 04 – 01]. https://www. openaire. eu/openaire-factsheet-for- repository-managers.
[17] Europe PMC [EB/OL]. [2018 – 03 – 05]. https://europepmc. org/.
[18] Arxiv. org[EB/OL]. [2018 – 03 – 10]. https://arxiv. org/.
[19] Dryad[EB/OL]. [2018 – 02 – 17]. https://datadryad. org/.
[20] Figshare[EB/OL]. [2018 – 04 – 05]. https://figshare. com/.
[21] CORDIS [EB/OL]. [2018 – 03 – 16]. https://cordis. europa. eu/.
[22] ORCID[EB/OL]. [2017 – 12 – 16]. https://orcid. org/.
[23] OpenDOAR[EB/OL]. [2018 – 02 – 20]. http://www. opendoar. org/.
[24] Re3data. org[EB/OL]. [2018 – 02 – 20]. https://www. re3data. org/.
[25] LA Referencia[EB/OL]. [2018 – 02 – 20]. http://www. lareferencia. info/joomla/en/.
[26] JAIRO[EB/OL]. [2018 – 01 – 15]. http://jairo. nii. ac. jp/.
[27] DOAJ[EB/OL]. [2018 – 01 – 21]. http://doaj. org/.
[28] DataCite[EB/OL]. [2018 – 04 – 02]. https://www. datacite. org/.
[29] OpenAIRE Guidelines[EB/OL]. [2018 – 01 – 02]. https://guidelines. openaire. eu/en/latest/.
[30] OpenAIRE guidelines for literature repositories[EB/OL]. [2018 – 01 – 02]. https://guidelines. openaire. eu/en/latest/literature/index. html.
[31] JEFFERY K, HOUSSOS N, JÖRG B, et al. Research information management: the CERIF approach[J]. International journal of metadata semantics & ontologies, 2014, 9(1):5 – 14.
[32] MANGHI P, HOUSSOS N, MIKULICIC M, et al. The data model of the OpenAIRE scientific communication e-infrastructure[C]//Metadata and semantics research. Berlin: Springer Berlin Heidelberg, 2012:168 – 180.
[33] CASTRO P D, SHEARER K, SUMMANN F. The gradual merging of repository and cris solutions to meet institutional research information management requirements[J]. Procedia computer science, 2014, 33:39 – 46.
[34] LORD P, MACDONALD A, LYON L, et al. From data deluge to data curation[EB/OL]. [2018 – 04 – 08]. http://www. ukoln. ac. uk/ukoln/staff/e. j. lyon/150. pdf.
[35] SCHIRRWAGEN J, MANGHI P, MANOLA N, et al. Data curation in the OpenAIRE scholarly communication infrastructure[J]. Information standards quarterly, 2013, 25(3):13 – 19.
[36] MANGHI P, BOLIKOWSKI L, MANOLD N, et al. OpenAIREplus: the European scholarly communication data infrastructure [J]. D-Lib magazine, 2012, 18(9/10):1 – 8.
[37] Guide for project coordinators in horizon 2020[EB/OL]. [2018 – 03 – 05]. https://www. openaire. eu/guide-for- project-coordinators-in-h2020.
[38] 吴建中. 推进开放数据 助力开放科学[J]. 图书馆杂志, 2018(2):4 – 10.
[39] 魏来, 李肸. 中日国家层面机构知识库服务比较研究——以 CHAIR 和 JAIRO 为例[J]. 情报资料工作, 2017(2):48 – 53.
[40] 崔海媛, 聂华, 罗鹏程,等. 资助机构开放获取知识库研究与构建 – 以国家自然科学基金基础研究知识库为例[J]. 图书情报工作, 2017,61(11):45 – 54.
[41] OpenAIRE detailed factsheet[EB/OL]. [2018 – 03 – 05]. https://www. openaire. eu/openaire- factsheetpdf.
[42] OpenAIREPlus detailed factsheet[EB/OL]. [2018 – 04 – 10]. https://www. openaire. eu/openaireplus-factsheet-pdf.
[43] The OpenAIRE2020 project[EB/OL]. [2018 – 03 – 01]. https://www. openaire. eu/project-factsheets.
[44] RCUK – Research Council UK[EB/OL]. [2018 – 05 – 08]. https://www. openaire. eu/rcuk-stats.
[45] SNSF – Swiss National Science Foundation[EB/OL]. [2018 – 05 – 07]. https://www. openaire. eu/snsf-stats.
[46] FWF – Austrian Science Fund[EB/OL]. [2018 – 05 – 04]. https://www. openaire. eu/fwf-stats.
[47] CARL announces collaboration with major European open science initiative, OpenAIRE[EB/OL]. [2018 – 05 – 12]. http://www. carl-abrc. ca/news/collaboration-with-openaire/.
[48] 陈淑珍, 韩珂, 祝忠明. 我国大学机构知识库建设的模式选择与实施策略[J]. 图书馆杂志, 2009, 28(8):52 – 54.
[49] MANOLA N, RETTBERG N, MANGHI P. OpenAIREplus project executive report[EB/OL]. [2018 – 02 – 08]. https://zenodo. org/record/15464#. Wt6o29KWaUk.
[50] 魏蕊, 王丽, 封勇丽,等. 京津冀高校机构知识库协同建设模式研究[J]. 图书情报工作, 2017,61(14):62 – 69.

Construction and Enlightenment of the OpenAIRE Project of the Open Access Infrastructure for Research in Europe

Sun Qian
Shaanxi Normal University Library, Xi'an 710119

Abstract: [**Purpose/significance**] This paper carries on research on the OpenAIRE project of open access infrastructure for research in Europe, analyzes its successful experiences, in hope of providing references to domestic open access repositories. [**Method/process**] Based on the methods of literature study and internet survey, this paper investigated and analysed the construction model and service function of OpenAIRE from organization model, type of data sources, interoperability guidelines, data curation, information retrieval function, statistics function and functions for multiple stakeholders. [**Result/conclusion**] The research summaries its successful experiences, including the open science content construction model which emphasizes the research data management and building a global collaborative open access repositories, the service functions which support open access policy and the management mode whose characteristic is from top to down, unified management, cooperation and coordination.

Keywords: open access repository　OpenAIRE　open access

《图书情报工作》2018 年优秀审稿专家

2018 年,有近 300 位审稿专家参加了《图书情报工作》稿件的同行评议工作,共评审稿件 1 400 余篇次,审阅 6 篇及以上的专家有 100 余位。高效、高质量的评审为《图书情报工作》遴选高质量稿件提供了保障。综合考虑今年以来的审稿数量、质量和时效,评选出 50 位 2018 年优秀审稿专家(名单如下)。《图书情报工作》为优秀审稿专家颁发证书并免费赠送一年期刊的电子版。感谢所有审稿专家对《图书情报工作》的大力支持!

(以下优秀审稿专家按姓名拼音排序):

姓名	工作单位
白如江	山东理工大学科技信息研究所
曹锦丹	吉林大学公共卫生学院
崔宇红	北京理工大学图书馆
邓胜利	武汉大学信息管理学院
邓小昭	西南大学计算机与信息科学学院
范爱红	清华大学图书馆
方向明	上海大学图书馆
冯　佳	上海社会科学院文学研究所
甘春梅	中山大学
高　凡	西南交通大学图书馆
韩　毅	西南大学计算机与信息科学学院
胡昌平	武汉大学信息资源研究中心
黄　崑	北京师范大学政府管理学院
黄国彬	北京师范大学政府管理学院
黄令贺	河北大学管理学院
姜春林	大连理工大学人文与社会科学学部科学学与科技管理研究所
李　刚	南京大学信息管理学院
李　明	南京大学信息管理学院
李　武	上海交通大学媒体与设计学院
李书宁	北京师范大学图书馆
李卓卓	苏州大学
刘　冰	天津师范大学管理学院
刘　勘	中南财经政法大学信息与安全工程学院
刘春丽	中国医科大学图书馆
刘晓娟	北京师范大学政府管理学院
刘雪立	新乡医学院期刊社/河南省科技期刊研究中心
刘兹恒	北京大学信息管理系
马　捷	吉林大学管理学院
马学良	国家图书馆
茆意宏	南京农业大学信息科学技术学院
牟冬梅	吉林大学公共卫生学院
裴　雷	南京大学信息管理学院
秦　鸿	电子科技大学图书馆
任树怀	上海外国语大学图书馆
邵　波	南京大学图书馆
滕广青	东北师范大学信息科学与技术学院
王立学	中国科学技术信息研究所
王晰巍	吉林大学管理学院
王延飞	北京大学信息管理系
吴　红	山东理工大学科技信息研究所
吴振新	中国科学院文献情报中心
向桂林	中国科学院生物物理研究所
闫　慧	中国人民大学信息资源管理学院
杨新涯	重庆大学图书馆
张广钦	北京大学信息管理系
张鹏翼	北京大学信息管理系
张卫东	吉林大学管理学院
赵　飞	北京大学图书馆
赵宇翔	南京理工大学
祝忠明	中国科学院兰州文献情报中心/中国科学院资源环境科学信息中心